Security Analysis of Biometric Template Protection

Koen Simoens, Stefaan Seys, and Bart Preneel
K.U.Leuven ESAT/SCD – COSIC
(Belgium)
NIST IBPC 2010

FP7 Integrated Project TURBINE (TrUsted Revocable Biometric IdeNtitiEs)

Background – TURBINE

- TURBINE project
  - TrUsted Revocable Biometric IdeNtitiEs
  - Privacy-enhanced solution for fingerprint biometrics
  - EU funded under FP7
  - http://www.turbine-project.eu
- Different protection methods developed by
  - Sagem Sécurité (France), Philips Research Europe (the Netherlands), Gjøvik University College (Norway)
- Evaluation tasks
  - Security testing: K.U.Leuven (Belgium – also legal evaluation)
  - Biometric performance testing: Gjøvik University College
- This talk reflects to some extent the security and privacy assessment of template protection techniques developed in TURBINE

Overview – Questions

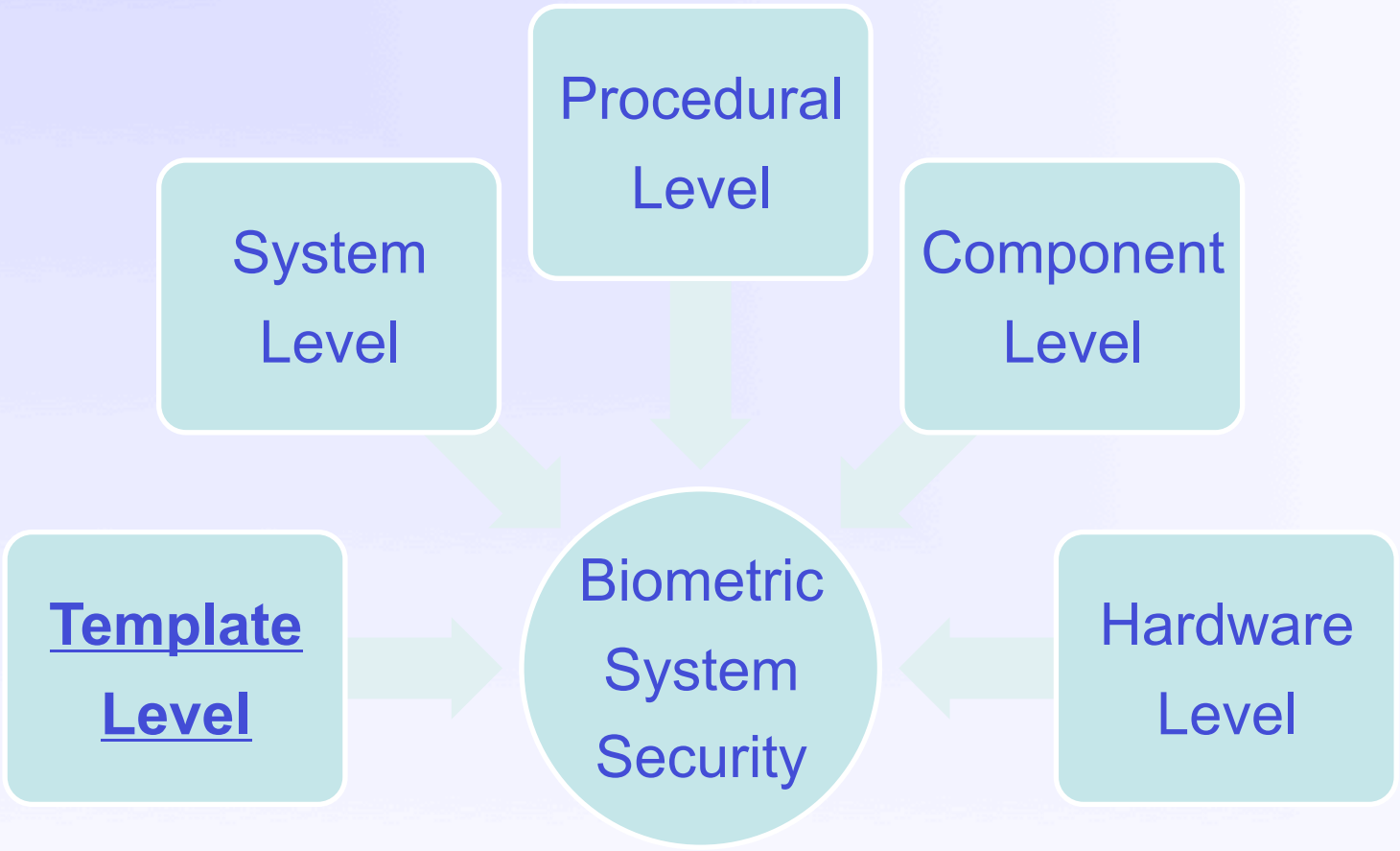
- What is biometric template protection?
- Why do we need it?
- What do we expect (requirements)?
- How can we achieve it (types)?
- **Which are the common pitfalls?**
- **Are their fundamental principles?**
- Where to start with the evaluation?
- How to compare results?
- What do we need more?

Overview

- The Meaning of Template Protection
- Why More Research on Template Protection is Needed
- Evaluation of Template Protection
- Objectives
  - Give intuition about fundamental principles
    - Design or analysis of new methods
  - Raise questions on how to evaluate
    - A common base for evaluation is needed

The Meaning of Template Protection

Biometric Template Protection at Multiple Levels



Template-Level Protection is Needed

- Many threats (impersonation, linking, etc.)
  - Conclusion: **do not store** reference data **in the clear**
- Current countermeasures
  - Encryption
  - Physical security
  - …

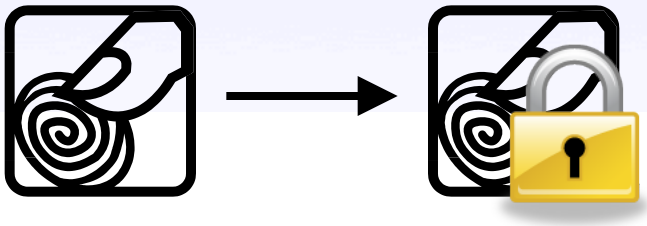
  => Complement with template-level protection
- Motivation from a risk management perspective (**what-if** analysis)
  - Physical protection may fail
  - Insider threats (trust assumptions no longer hold)
  - Desired **renewability** feature

Protection at Template Level

- **Biometrics-only** model: “cannot be lost/forgotten”
  - Assume no keys, passwords or smart cards for security
  - Possibly token as storage medium
  - Biometrics are **secrets**, but they are **noisy**
    - Classical data privacy schemes do not work
- Different methods have been proposed:
  - Quantization schemes
  - Discrete schemes
  - Mixed quantization/discrete schemes
  - Cancelable biometrics
  - …

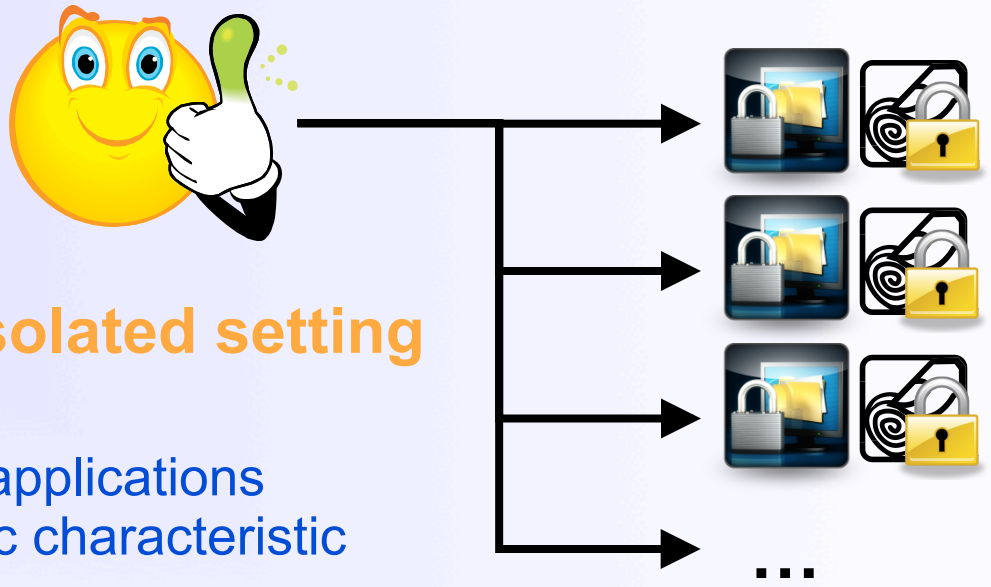
Template Protection in TURBINE

- Fuzzy schemes allowing noisy inputs
  - **Public data** helps to reliably extract bits
  - Hash extracted bits and store as reference
  - **Error-correction** to deal with noise
  - Implicit comparison: wrong result if distance > t
  - Classic example:
    - Fuzzy commitment (Juels & Wattenberg, CCS '99)
- TURBINE **pseudo identity** model
  - ISO 24745 Biometric Template Protection



Auxiliary Data (AD)
+
Pseudo ID
(secure reference)

Multi-Application Scenario



- Schemes **secure in isolated setting**
  - But reality…
  - Registered in multiple applications with the same biometric characteristic
- Multiple databases with protected templates
- We tend to forget
  - Different applications = different algorithms

Properties

- Main goals of template protection
  - One-way transformation: **irreversibility**
  - Diversification: **unlinkability** and revocability
  - Maintain biometric performance!
- Subtle issues
  - What does reversibility mean?
    - Reverse to enrolment sample or to other genuine/ impostor sample
  - Two-template irreversibility

Why More Research on Template Protection is Needed

"Fundamental" principles

A Word of Caution

- Personal ideas/opinion
- Not all principles are yet fundamental
  - Unproven, but give intuition
- Valid for all types of template protection?
- Recall
  - Template-level protection complements protection at other levels

User-Specific Side Information

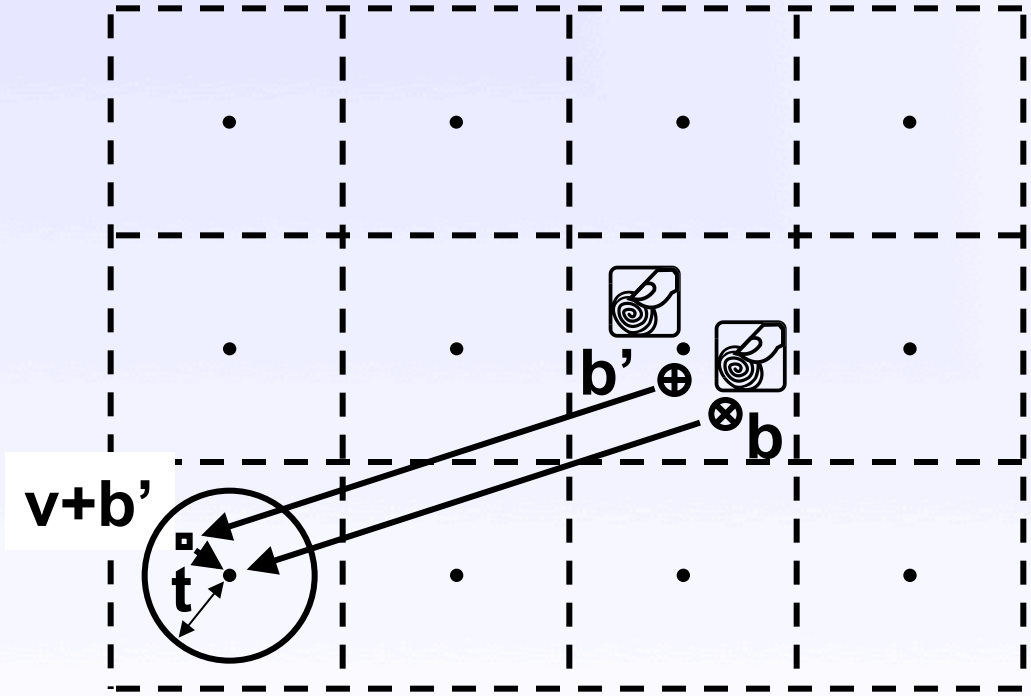
- Isometric **one-way** transformation
  - If b is a biometric sample and {b'} its neighbours
  - Take b somewhere else, thus {b'} also
  - This is why cryptographic hashing doesn't work
- There is no single transformation for all
  - Transformation is adjusted to enrolment sample
  - Side information **depends on input**
    - User-specific auxiliary data (public helper data)

Leakage Is Unavoidable

- What is leakage?
  - Information that reduces “uncertainty” about the enrolment sample
  - *It becomes “easier to guess”*
  - **Entropy reduction**
    - Fuzzy extractors (Dodis et al. EUROCRYPT 2004)
    - Adam Smith (Ph.D. Thesis 2004)
    - Only for discrete sources
- Where does it come from?
  - It is in the side-information
  - It is needed to **compensate noise**
  - Leakage is tolerated but should not be ignored

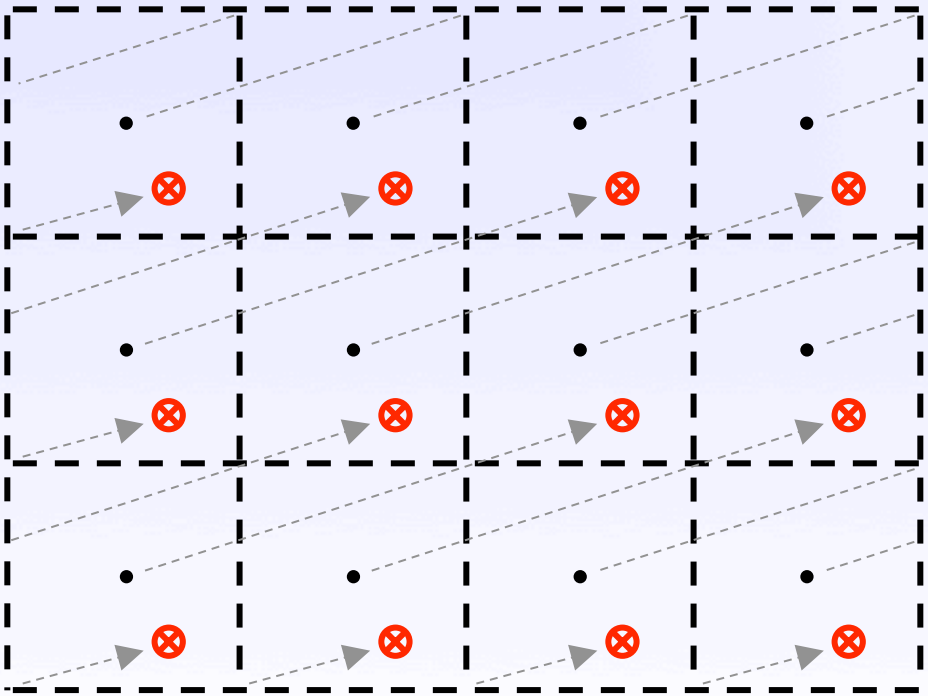
Leakage Is Unavoidable – Example

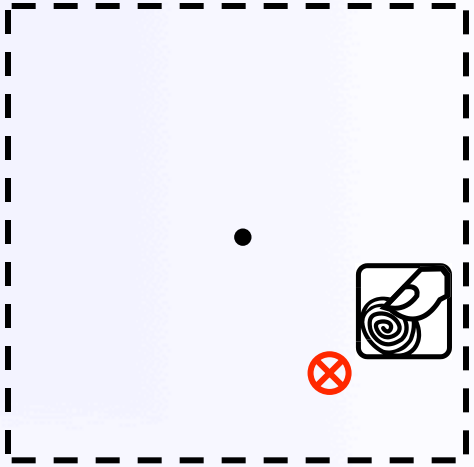
- Fuzzy commitment
  - Enrolment sample w, probe w'
  - Offset v (translation preserves distance)
  - Decode v + b' to c' and verify if h(c') = h(c)



Leakage Is Unavoidable – Example

- Theoretically proven **AD must leak** information
- Inverse code-offset from any codeword
- Uncertainty is reduced (no actual bits leaked)
- **Position in square is revealed** implicitly



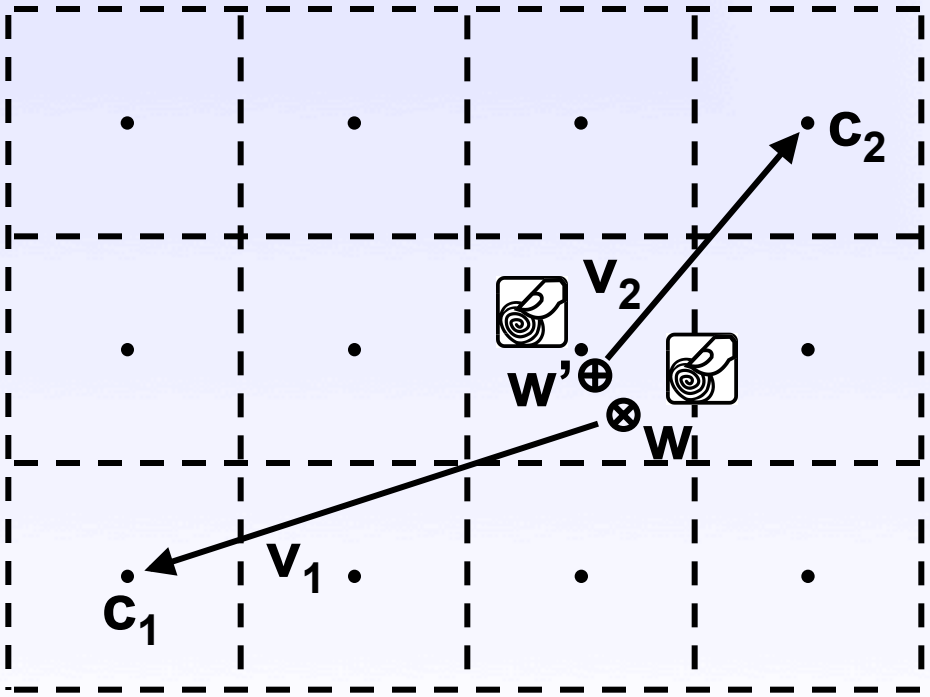


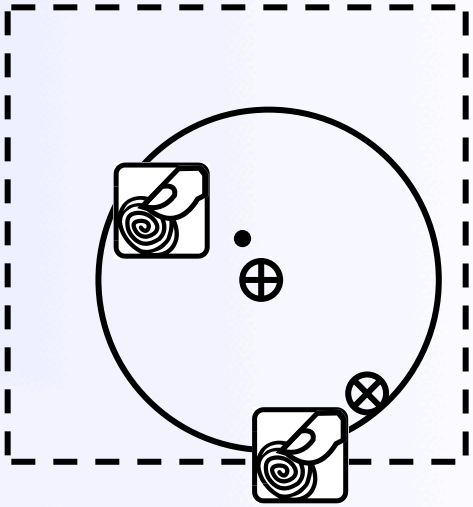
We Leak Too Much

- Current schemes leak to much
  - More leakage than needed to correct errors
  - This is why **cross-matching works** for fuzzy commitment and some quantization schemes
    - Simoens et al. S&P 2009
    - Buhan et al. SPEED 2009
- Can we improve?
  - Mathematical bounds (coding theory)

We Leak Too Much – Example

- Completely different PIs and ADs
- Cross-matching based on relative positions
- **Successful attack** against fuzzy commitment



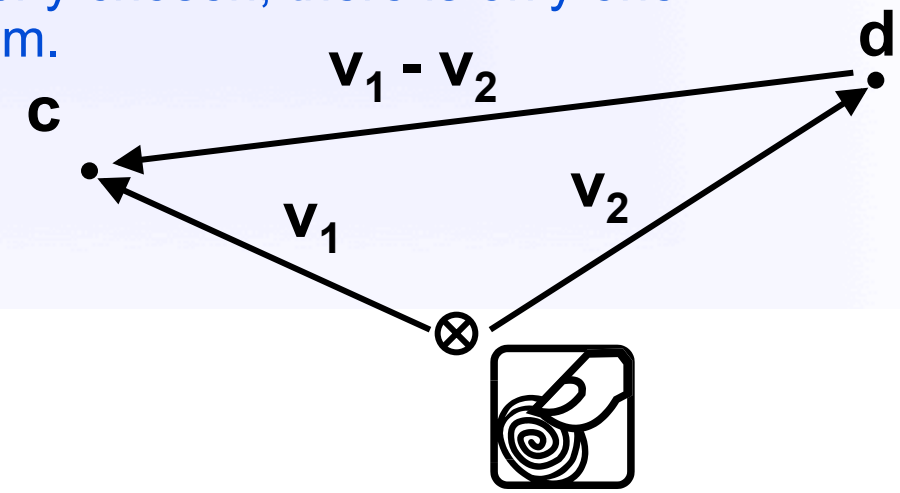


If You Leak, Leak Consistently

- Leakage is unavoidable
- **Leak the same** in different applications
  - This implies using the same algorithm
  - Impossible to maintain in practice?
- If not, **reverse two protected templates**
  - Theoretical attack (Simoens et al. S&P 2009)
  - Easy to see for discrete biometrics
    - Code-offsets, projection based
  - What about cancelable biometrics?

If You Leak, Leak Consistently

- Assume exactly same input is used
  - In practice enrolment samples are not equal
- Codewords c and d are from **different codes**
- Subtract offsets to obtain v = v1 − v2 = c − d
- Solve mathematical problem
  - Find c and d, from the first and the second code, who's difference equals v
  - If codes are not properly chosen, there is only one solution to this problem.



Linear = BAD

- Linear
  - "matrix", "offset", "translate", "rotate", "XOR"(!)
  - **Is not good** (cf. cryptography)
- Impact
  - Code-offsets
    - **Linkability** (cross-matching)
    - **Two-template reversibility**
  - Transformation of point-based features (e.g. minutiae sets)
    - **Correlation** between minutiae is preserved
- Non-linearity
  - Where to get it?
  - Conflicts with isometric input transformation

Other Observations

- **FAR attacks** are inherent to biometrics
  - Template-level protection requires additional measures
  - FAR attacks hurt more than you'd expect
    - Recover enrolment data from false accept
- Some schemes allow a **wider input range**
  - Examples: projection-based schemes, superimposing minutiae subsets
  - Attempt to counter some of the above attacks
  - Curiosity in Template Protection
    - Theoretical increase in false acceptance rate?
- Theory differs from practice
  - Evaluation requires working with real data
  - Distribution of binarized templates matters

Other Observations

- Including **hardware is out of scope** in the biometrics-only model
  - But it works! At least, we think so…
  - Nicer properties
- Information content (**entropy**) **is limited**
  - How many minutiae in a fingerprint?
  - What is the scanner resolution?
  - What is the range of the coordinates?
  - If you take into account noise tolerance?
- Is it not all discrete?
  - Quantization schemes
  - In the end… all the same?


TURBINE

Evaluation


SEVENTH FRAMEWORK
PROGRAMME


KATHOLIEKE UNIVERSITEIT
LEUVEN

Before You Start Evaluating

- Do you have a proper **framework**
  - To model/analyze ALL methods
    - E.g. fuzzy extractors for discrete biometrics
  - To set proper terms of reference
    - How to define and measure security properties?
- What is it that you want?
  - Application requirements
  - Are you willing to trade between properties
    - E.g. irreversibility vs. unlinkability
  - Are the requirements realistic and needed in practice?
- Get into the right mindset
  - Become a non-believer

During Evaluation

- Set clear adversary **objectives**
  - E.g. we want to compare/link protected templates
- **Test principles** mentioned above
  - Where is the leaked information?
  - In which form is it leaked?
  - How much is leaked and can we use it?
- How to **rank** different protection methods
  - Difficult without a unified framework
  - Already some consensus on security properties
  - Ongoing discussion in TURBINE and elsewhere

Expectations for The Future

- We need more **advanced techniques**
  - Non-linear methods
- We need more **formal approaches**
  - Unified security notions
  - Less “we think/believe…”
- We need **provable security**
  - Cf. public-key cryptography: security proven under some number-theoretic assumptions
- Are we ready for the first Template(-level) Protection Standard?

Thanks

koen.simoens@esat.kuleuven.be

http://www.turbine-project.eu

This work is supported by funding under the Seventh Framework Programme of the European Union, Project TURBINE (ICT-2007-216339). All information is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. The European Commission has no liability in respect of this document, which is merely representing the authors' view.

References

- Buhan, I., Breebaart, J., Guajardo, J., de Groot, K., Kelkboom, E. and, Akkermans, T. 2009. **A quantitative analysis of crossmatching resilience for a continuous-domain biometric encryption technique**. In Proc. Of the First International Workshop on Signal Processing in the EncryptEd Domain (SPEED 2009).
- Dodis, Y., Ostrovsky, R., Reyzin, L., and Smith, A. 2008. **Fuzzy extractors: How to generate strong keys from biometrics and other noisy data**. SIAM J. Comput. 38, 1, 97-139. http://dx.doi.org/10.1137/060651380
- Juels, A. and Wattenberg, M. 1999. **A fuzzy commitment scheme**. In Proc. of the 6th ACM Conference on Computer and Communications Security - CCS '99, 28-36. http://doi.acm.org/10.1145/319709.319714
- Simoens, K., Tuyls, P., and Preneel, B. 2009. **Privacy weaknesses in biometric sketches**. In Proc. of the 2009 30th IEEE Symposium on Security and Privacy, 188-203. https://www.cosic.esat.kuleuven.be/publications/article-1205.pdf
- Smith, A. D. 2004 **Maintaining Secrecy when Information Leakage is Unavoidable**. Doctoral Thesis. Massachusetts Institute of Technology.